Meaningful Integration of Data, Analytics and Services

Grant Agreement No. 727721

Contract Duration: 40 months (1st November 2016 – 29th February 2020)

This project is funded by
The European Union

H2020-SC1-2016-CNECT
SC1-PM-18-2016 - Big Data Supporting Public Health Policies

Deliverable 2.4

Report on Public Perceptions of Identifiable Data

Circulation:	PU (Public)
Nature:	R (Document/Report)
Version #:	2.2
Issue Date:	04/06/2019
Responsible Partner(s):	DH PHE
Author(s):	Ava Hodson, Natasha L Bloodworth, & Dale Weston
Status:	Final
Reviewed on:	22/07/2019
Reviewed by:	MIDAS Executive Board
Contractual Date of Delivery:	31/10/2018

Executive Board Document Sign Off

Role	Partner	Signature	Date
Project Coordinator	Ulster	Michaela Black	28-06-19
WP1 Lead	Ulster	Michaela Black	28-06-19
WP2 Lead	SET	Paul Carlin	15/07/2019
WP3 Lead	VICOM	Gorka Epelde	29-06-19
WP4 Lead	KU Leuven	Gorana Nikolic	10-07-19
WP5 Lead	VTT	Juha Pajula	04-07-19
WP6 Lead	DCU	Regina Connolly	30-06-19
WP7 Lead	Ulster	Jonathan Wallace	19-07-19
WP8 Lead	Ulster	Michaela Black	28-06-19
Scientific-Technical Manager	Analytics Eng	Scott Fischaber	17-07-19

Abstract

Personal data has potential uses within a range of public health management, illness monitoring and policy development contexts. Members of the public may be willing to share their information based on its perceived utility for health administration or protection. However, the perception of stigma, risk/threat and the level of identifiability from data may alter this. Participants were shown a range of different scenarios that varied in terms of whether they presented a high or low risk/threat scenario, a high or low stigma scenario, and asked participants to consider sharing one of three different types of data(anonymous vs. linked-identifiable vs identifiable). All participants responded to all different scenario combinations. 137 individuals participated in the survey, but only 57 (43.2%) fully completed the survey. It was found that data type has a key and consistent effect on perceptions of data sharing: individuals are happier to share their data to the extent that it is anonymised rather than identifiable. Risk/threat and stigma also exert an influence on perceptions of/ attitudes towards data sharing, however these findings were inconsistent.

Copyright

- Ulster – University of Ulster (Project Coordinator) (UK)
- DCU – Dublin City University (Ireland)
- KU Leuven – Katholieke Universiteit Leuven (Belgium)
- VICOM – Fundación Centro De Tecnologías De Interacción Visual y Comunicaciones Vicomtech (Spain)
- UOULU – Oulun Yliopisto (University of Oulu) (Finland)
- ANALYTICS ENG – Analytics Engines Limited (UK)
- QUIN – Quintelligence D.O.O. (Slovenia)
- BSO – Regional Business Services Organisation (UK)
- DH – Department of Health (Public Health England) (UK)
- BIOEF – Fundación Vasca De Innovación E Investigación Sanitarias (Spain)
- VTT – Teknologian Tutkimuskeskus VTT Oy (Technical Research Centre of Finland Ltd.) (Finland)
- THL – Terveyden ja hyvinvoinnin laitos (National Institute for Health and Welfare) (Finland)
- SET – South Eastern Health & Social Care Trust (UK)
- IBM Ireland Ltd – IBM Ireland Limited (Ireland)
- ASU ABOR – Arizona State University (USA)

The MIDAS project is funded under the EC Horizon 2020 SC1- PMF-18 Big Data Supporting Public Health Policies

Document History

Version	Issue Date	Stage	Content and Changes
1.0	26/03/2019	Draft	Draft of report uploaded
2.0	04/06/2019	Draft	Comments from SET incorporated, cosmetic edits to Introduction, Method, existing Results, and Discussion. Changed formatting of Table 2 to fit with rest Additional supplementary analysis prepared (section 5.2), Appendix 4 removed (as supplanted by additional analysis), amendments to section 6.4 in light of additional analysis. Minor formatting amendments throughout
2.1	03/07/2019	Draft	Comments and edits by Ulster and VICOM actioned and incorporated.
2.2	16/07/2019	Draft	Further comments and edits actioned
2.3	17.07.2019	Draft	Fixed notation in main text Table 1, accepted typographical changes, amended paragraph spacing.

Statement of Originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Executive Summary

Work Package:	Work Package 2: Governance, Consent, and Privacy
Work Package leader:	SET
Task:	T 2.3 – Implement New Model
Task leader:	SET

Personal data has potential uses within a range of public health management, illness monitoring and policy development. Members of the public may be willing to share their information based on the utility of their personal data, for purposes of health administration or protection; for example, as an operational resource in the aftermath of a major incident. However, the perception of stigma, risk/threat and the level of identifiability from data may alter this.

This study involved examining participants reactions to a variety of different scenarios that varied in terms of the risk/threat presented (either high risk/threat or low risk/threat), the stigma presented (either high stigma or low stigma) and the type of data that participants were asked questions about (either anonymous data, linked identifiable data, or identifiable data). For example, one scenario involved participants considering a situation where they had witnessed a terrorist attack (high threat), and had subsequently been drinking a lot (alcoholism - high stigma), and asked participants to consider whether they would be willing to share identifiable data. All participants received all combinations of these types of scenario (all types of risk/threat, all types of stigma, all types of data) and responded to questions about each. In statistical terms, this is a three factor design (risk/threat, stigma, and data type) with two levels for risk/threat (*high vs low*), 2 levels for stigma(*high vs low*), and 3 levels for data type (*anonymous vs. linked-identifiable vs identifiable*); in other words a 2x2x3 within-subjects design.

137 individuals participated in an online survey, but only 57 (43.2%) fully completed the questionnaire. All participants were shown four vignettes reflecting high vs low stigma (alcoholism vs. physical injury) and high vs low threat/risk (terrorist incident vs routine health problem). Participants were subsequently asked a series of questions regarding their perceptions of, and attitudes towards, the sharing of three different kinds of data (anonymous, linked-identifiable, and identifiable).

Considered together, the results suggest that individuals are happier to share their data when it was associated with increased privacy. In other words, perceptions towards data sharing were more positive to the extent that the data was anonymous, rather than linked-identifiable (described as data which is ostensibly anonymous but may be linked to other pools of information, thus removing anonymity), or identifiable data.

The effects of stigma and risk/threat on these outcomes were less consistent. Risk/threat had a main effect on outcomes, wherein high risk inferred greater urgency of sharing data, potentially guided by the notion of being ‘for the greater good’ (Dogan, 2015; Rubin et al., 2018). Individuals were less happy to share information under conditions of high stigma, particularly where it was

a) without consent or
b) with organisations that are not typical healthcare organisations, such as other non-healthcare government and private healthcare providers.

Lastly, preliminary additional analysis reveals that there are additional differences in perceptions dependent upon the data recipient (i.e., healthcare provider vs. private company), which should be scrutinised further in future research.

Table of Contents

1 Introduction

Personal data has many potential uses within public health management, illness monitoring and policy development (McGrath, 2004; Taylor, 2010; Jones & Gagneja, 2016). Since data-linking enables the collation of large sets of personal information, these platforms could be used for online, epidemiological data management or during mass emergencies (Kostkova, 2018; Rubin et al., 2018). However, data-linking in such a way increases the chances of being identified, which is a main worry for members of the public (Medical Research Council [MRC], 2007; Dogan, 2015).

In addition to worries about identification, individuals are influenced by the perception of stigma and risk/threat in their acceptability of data-sharing. For example, the social stigma associated with certain health data, like mental illness, may influence individuals to be less willing to share this type of information, especially when compared to data concerning physical ailment (Schomerus et al., 2010). However, events that are high in risk produce distinct emotional and behavioural responses which may counteract this unwillingness (Gigerenzer, 2006).

With few exceptions, however, these constructs have not been extensively explored in the context of public responses to the sharing of personal data. Whiddett and colleagues (2006) identify that there is a lack of research from the patients' perspective. This study attempts to fill this gap. Whilst there is a relative dearth in research in this area, it is important to consider the relevant streams of literature which do consider the impact of context, data type, stigma and perceived risk/threat on behavioural outcomes and perceptions of data sharing.

1.1 Data sharing and perceptions of privacy

Public health management has become increasingly dependent upon effective data use, whereby mobile access to the internet is promoting nuanced delivery of healthcare and well-being interventions (Whiddett, Hunter, Engelbrecht, & Handy, 2006; Thackers, Qualter, & Lee, 2012; Kostkova, 2015). Accordingly, the 'information age' has enabled the collation of large and complex datasets that integrate data from traditional forms of health monitoring, as well as other indirect determinants of health, such as environmental factors and social behaviour (AbouZahr, Adjel, & Kanchanachitra, 2007; van Panhuis et al., 2014).

Linking all forms of personal information allows for the compilation of structured and valuable data-sets, with utility for epidemiological management and surveillance (Chew & Eysenbach, 2010; Poikola, Kuikkaniemi, & Honko, 2014; Kostkova, 2018). However, linking disparate, ostensibly anonymous datasets (henceforth referred to as *linked-identifiable* data) makes individuals potentially identifiable and this concept of 'data-linking' may contribute to the concerns of the public.

Pseudonymisation aims to protect personal information in a way that it can no longer be directly linked to a particular individual (Ewerlöf, 2018); this process involves the removal of identifying information, like name or date of birth, such that individuals may be more willing to share this type of information (than when fully identifiable), given that their identity is still covered. However, when this action is breached it evokes concerns about a progression towards Orwellian surveillance and increases identifiability, both being an invasion of privacy which may limit an individual's wish to share data (McGrath, 2004; Taylor, 2010; Wellcome Trust, 2015; Jones & Gagneja, 2016).

In particular, issues have been highlighted regarding the sharing of directly identifiable, health-related information, whereby it represents the breakdown of confidentiality which underpins the clinician-patient relationship. More elaborate or sophisticated use of personal information could erode personal privacy and individuals may lose autonomy over how and under what circumstances their data is used (Whiddett et al., 2006; AbouZahr et al., 2007; Boerma & Stansfield, 2007; Murray, 2007; Chan et al., 2010; Taylor, 2010; van Panhuis et al., 2014).

This issue has come under further intense scrutiny since the Cambridge Analytica scandal; a click-bait [1] quiz, which enabled Cambridge University to gain access to the personal data of more than 50 million Facebook profiles (Granville, 2018; Landau, 2018). This information was subsequently used to influence the results of the 2016 Presidential Election and EU referendum (Granville, 2018; Landau, 2018). Following this, another two data-breaches via social media platforms were reported, where over 50 million users had their data accessed and sold (Wong, 2018; Zakharov, 2018).

---

[1] Clickbait is defined as internet content, the main purpose of which is to attract attention and encourage visitors to click on a link to a particular web page (Clickbait, 2019).

As such, governance and regulation frameworks for data privacy and data use are perceived to be "lagging behind" the potential for health technology and effective data use (van Panhuis et al., 2014; Kostkova, 2018, p.1). Prompted by this, the strengthened General Data Protection Regulation (GDPR) was created to replace the outdated regulations of the Data Protection Act (1998) and included guidelines for pseudonymisation of personal information (Varotto, 2015; Astrup, 2018; Ewerlöf, 2018).

Public apprehension towards data-sharing (either directly identifiable or potentially identifiable via data linkage) is therefore not unfounded and contributes to a sense of general 'distrust' in how data is feasibly used (van Panhuis et al., 2014). In addition, whilst allowing one's data to be included in aggregate datasets is deemed as 'valuable', there are public concerns about data-breaches, accidentally making personal identifiable data (PID) public, and how it would then be used (Kaufman, Murphy-Bollinger, Scott, & Hudson, 2009; Rubin et al., 2018). Protecting personal information and ensuring anonymity are therefore paramount for ensuring public confidence in data sharing (Barrett et al., 2006).

To explore the impact of data type on attitudes and perceptions around data sharing, the present study asks participants a range of questions related to the sharing of different types of data (anonymous, linked-identifiable, and identifiable) across a range of different health and security related contexts. In doing this, we hope to elucidate the general public's perceptions of sharing these forms of data in the context of a 'post-Cambridge Analytica' and 'GDPR era'.

The acceptability of sharing personal data is believed to be fundamentally influenced by who the recipient is. Over 50% of respondents to a survey investigating patients' attitudes towards sharing health information, were not willing to share their personal data with private health insurers or government agencies; on the other hand, there was a greater perceived acceptability of sharing data with health professionals (Barrett, Cassell, Peacock, & Coleman, 2006; Whiddett et al., 2006). Indeed, although the public are 'somewhat' or 'very concerned' about the privacy of their medical information (Kaufman et al., 2009), in the context of wellbeing following a major incident, participants expected their General Practitioner (GP) to be notified and linked to their personal medical records (Dogan, 2015).

This may be related to the notion that, individuals are more willing to share PID when it is for a wider public health benefit or related to relevant public health issues, such as an invitation to a cancer screening. For example, Barrett and colleagues found that participants did not consider knowledge of name or address, within the National Cancer Registry, to be an ‘invasion of privacy’ (Barrett et al., 2006). The perceived public benefit of data-sharing under these circumstances may therefore increase public acceptability (Kaufman et al., 2009; Rubin et al., 2018). Thus, individuals may wish for more sensitive information to remain private but would be most willing to share this type of data with a doctor or practice nurse, with emergency personnel second to these (Whiddett et al., 2006; Rubin et al., 2018).

In addition, members of the public may be more willing to share their personal data if it were to be more closely linked to wider health-related administration, for example booking appointments or ordering prescriptions (Keahey, 2018). Accordingly, the acceptability of data sharing may be wholly dependent on context; since ‘data-sharing context’ and the relationship between this and data type (i.e., identifiable or anonymised) has not been empirically tested (Ipsos, 2014; Rubin et al., 2018), the present study will provide valuable insights into the acceptability of data-sharing.

1.3 Risk/threat perception

After the 7/7 London bombings, the cautious approach to sharing personal information of victims hampered survivors’ access to support services and networks, leading to a recommendation to share PID (HM Government, 2007). Rubin and colleagues (2018) suggest that the use of PID datasets, as well as open and rapid data sharing, may be efficacious in mass emergencies, especially for those individuals who may require further psychological or physical aid. Freely available data is a potentially important operational resource for emergency responders and victims alike, to both facilitate emergency response and to help reduce the impact of maladaptive behaviour (Rubin et al., 2018).

As aforementioned, research has found that individuals may be more willing to share PID following a major incident (Dogan, 2015). This willingness may be informed by the individuals’ perception of risk or threat surrounding these types of events; although mass emergencies are low-probability, they can be high-damage and thus elicit distinct emotional and behavioural responses (Gigerenzer, 2006). Unfamiliarity associated with adverse incidents may be particularly anxiety-evoking, which infers a greater perception of risk or potential threat (Slovic, 1987; Siegrist et al., 2007; Brown, 2014). Their emotional significance influences behavioural outcomes, such as avoidance, to compensate for related anxiety (Gigerenzer, 2006). This could infer an avoidance to share personal information after major incidents. Indeed, the perceptions of risk associated with a given event have previously been associated with willingness to share personal information; Rubin and colleagues (2018) found that significantly more individuals believed personal data should be shared following a radiation-related incident, including with a general practitioner, than following a mass-shooting incident.

Considering this literature together suggests that novel threats that are high impact and low probability – like terror incidents – may therefore change an individual's acceptability of data sharing. As such, individuals may be more reluctant to share data under normal circumstances, especially when compared to data sharing that may be related to a novel or dangerous event. Despite this possibility, to the best of the authors' knowledge there is little literature (other than that cited herein) surrounding the public acceptability of data-sharing in public health emergencies and the consequent access to relevant and important data, amidst this type of scenario. The present study aims to investigate whether varying level of risk/threat within scenarios, will impact on the willingness to share data.

1.4 Social stigma

Social stigma represents another potential factor that might influence individuals' willingness to share their data. The notion of stigma is applied to a range of conditions, from social status to job-type to physical ailments, chronic illness or mental illness (Fife & Wright 2000; Phelan, Link, Stueve & Pescosolido, 2000; Link & Phelan, 2001; Logie & Gadalla, 2009; O'Donnell, Corrigan & Gallagher, 2015). "Labelling, stereotyping … and discrimination" maintain social stigma through the [mis]attribution of particular behaviours or expectations to certain social groups (Link & Phelan, 2001, p. 367; Corrigan, 2004). As such, certain health conditions bear particular social significance and are associated with specific labels or social identities (Link & Phelan, 2006; O'Donnell et al., 2015). For example, although alcohol dependency is classified as a mental illness, it is more commonly attached to the stereotypes of unpredictability and danger (Schomerus et al., 2010).

When negative, social stigma becomes social prejudice and is accompanied by "profound ... effects on social behaviour" (Schomerus et al., 2010, p. 105). When this process of labelling is applied to health practices, it highlights why individuals may be reluctant to receive psychiatric diagnoses or engage in treatments (Klingemann, 2001; Corrigan, 2004; Link & Phelan, 2006; Fear, Seddon, Jones, Greenberg & Wessely, 2012; Wallhed Finn, Bakshi & Andreasson, 2014). For example, if individuals are associated with conditions they believe to be socially stigmatised (such as, mental health conditions) they may be less willing to share personal information altogether. As such, the connotations of certain diagnoses may influence public acceptability of sharing personal, illness-related data.

2. Rationale

Although personal data has been identified as potentially useful resource within healthcare and in the aftermath of disasters (McGrath, 2004; Taylor, 2010; Jones & Gagneja, 2016; Rubin et al., 2018), there is a relative dearth of information regarding the public acceptability of data-sharing for these purposes. The present study sought to address this lacuna by addressing how acceptable members of the public find the sharing of different types of data, with particular focus on exploring the impact of data type, stigma, and perceived risk/threat. Specifically, the present study is a vignette-based piece of research in which participants read four different health and security related scenarios (designed to elicit differing levels of stigma and perceived risk/threat in respondents) and discuss their perceptions of and attitudes towards sharing different types of data (anonymous, linked-identifiable, identifiable). The results are subsequently discussed in relation to the literature discussed herein.

3. Hypotheses

Based on the literature reviewed above, it is hypothesised that there will be a main effect of data-type on participants’ perceived acceptability of data sharing (Whiddett et al., 2006; Ewerlöf, 2018). Specifically, participants are expected to be more willing to share anonymous data than either linked-identifiable or identifiable data, and more willing to share linked-identifiable than identifiable.

Main effects of stigma and risk/threat are also hypothesised, such that participants will be more positive about sharing data related to low-stigma scenarios and high-risk/threat scenarios, than data related to high stigma scenarios and low risk/threat scenarios. These hypotheses recognise the operational potential of personal data after major incidents, but the reluctance to share data that is highly stigmatised (Corrigan, 2004; Link & Phelan, 2006; Rubin et al., 2018).

Furthermore, this hypothesised effect of stigma on acceptability of data sharing may also be attenuated by the perceived risk/threat associated with the given scenario (an interaction effect). Specifically, given the previously discussed impact of high-risk/threat contexts (e.g., terrorist attacks) on people's willingness to share data, it may be that the previously hypothesised main effect of stigma will only hold under conditions of low risk/threat. The difference in acceptability of sharing data related to stigmatised and non-stigmatised conditions would be attenuated under conditions of high risk/threat, where the data sharing is perceived as being for the 'greater good'. Finally, given the relationship between social stigma and identification, the expected interaction between stigma and perceived risk/threat may only hold for anonymous data (a three-way interaction effect).

4. Method

4.1 Participants & Design

Data collection took place between September 2018 and December 2018. In total, 137 individuals participated in the survey, but only 57 (43.2%) completed the full questionnaire. Of these, 75.4% were female, 79% identified themselves as British or Irish, and had a mean age of between 25 and 34 years. 93% had degree-level qualifications (including Bachelors, Masters and Doctoral degree). One participant indicated that they did not have access to the internet, despite completing this online survey. Only the 57 participants who had fully completed the survey were included in the final data analysis, this enabled us to minimise any impact of individuals repeatedly completing the survey (as it was deemed unlikely that individuals would fully complete the survey multiple times, perhaps just starting again if they had gotten cut off part way through, for instance). Indeed, the demographic data for all 57 included participants were visually scanned by two researchers, revealing no obvious duplicate respondents. Values of 6 – relating to 'not applicable' in survey responses – were re-coded as missing data and were thus excluded from analysis.

This study involved examining participants reactions to a variety of different scenarios that varied in terms of the risk/threat presented (either high risk/threat or low risk/threat), the stigma presented (either high stigma or low stigma) and the type of data that participants were asked questions about (either anonymous data, linked identifiable data, or identifiable data). For example, one scenario involved participants considering a situation where they had witnessed a terrorist attack (high threat), and had subsequently been drinking a lot (alcoholism - high stigma), and asked participants to consider whether they would be willing to share identifiable data. All participants received four scenarios combining these different elements (all types of risk/threat, all types of stigma, all types of data) and responded to questions about each (more detail is presented in the next section). These materials are discussed in the following subsection.

Ethical approval for this study was granted by the Public Health England Research Ethics and Governance Group (reference number: R&D358) and was subsequently approved by the MIDAS Project Ethical and Privacy Advisory Group (request ID: 15).

4.2 Materials & Procedure

Participants were recruited using online invitations distributed by University and consortium partners, and via social media sites such as Twitter. Potential participants were provided with a link to the information sheet and consent form for participation. Following the provision of consent, all participants answered a series of demographic questions (age, gender, nationality, education, employment status, occupation, whether or not the individual considered themselves to have a disability, country of residence, and internet use).

Following this, all participants were shown each of four vignettes, in a counterbalanced order based on a Latin square design, to reduce the impact of order effects (i.e., different participants read and responded to the scenarios in different orders). The vignettes were based on case studies developed by HM Government (2007), adapted to allow for the manipulation of stigma and risk/threat. The final four vignettes were drawn from a larger pool of potential scenarios that were pilot tested for both stigma and risk/threat content (see Appendix 1 for results of pilot testing). In each scenario, participants were asked to imagine that they:

1. Have been witness to a terror attack, sustain no physical injuries but are referred to a specialist team because of the amount of alcohol they are drinking because of the trauma;
2. Have been witness to a terror attack, during which they fall and sustain a fractured ankle;
3. Are concerned about the amount of alcohol that they are drinking, and referred to a specialist team;
4. Have fallen down the stairs and fractured their ankle.

Dread risks/threats, that are low probability, high-consequence events, were represented within the present study by ‘witnessing a terrorist incident’ (scenarios 1 and 2); low risk/threat were more routine conditions requiring a GP visit (scenarios 3 and 4). Stigma was manipulated through the experience of physical (a broken bone; scenarios 2 and 4) versus mental illness (alcohol dependence; scenarios 1 and 3), with the view that mental illnesses are more highly stigmatised than physical injuries (Schomerus et al., 2010). Table 1 presents a graphical representation of how the scenarios relate to the manipulated variables. See Appendix 2 for the final scenarios.

Table 1. *Details of each scenario in terms of level of stigma and risk/threat.*

		Risk/threat	
		High	Low
Stigma	High	1	3
	Low	2	4

Following each scenario, participants completed the same questionnaire items. Each participant therefore completed four versions of the same questionnaire, one after each scenario.

Another outcome for consideration is the organisation with which information is being shared. This is based on the notion that public acceptability differs dependent on who the data-recipient is, varying from public to private companies (Barrett, Cassell, Peacock, & Coleman, 2006; Whiddett et al., 2006). For example, private companies may be viewed more negatively, compared to those which are ostensibly health-related, such as the NHS, Public Health England or personal general practitioners.

Based on research conducted by Whiddett and colleagues (2006), different data recipients will be presented in this study. Participants will be asked about their willingness/ happiness for data to be shared with doctor, university, healthcare organisation (e.g. NHS), healthcare related government organisation (e.g. Public Health England [PHE]), a non-healthcare related government organisation, an independent charity (e.g. Cancer Research), or a private medical research company. The same questions were asked related to anonymous, linked-identifiable and potentially identifiable information (*see* Table 2).

Table 2. *Details of questionnaire items.*

How necessary do you think it is to share this information?
How urgent would you consider the sharing of this information to be?

I would be happy for my data to be shared
I would be happy for my data to be shared if my consent were requested from my doctor/other healthcare professional
I would be happy for my data to be shared without my explicit consent
I would be happy for my data to be shared with the following organisations...
...My doctor
...Academic researchers from a University
...A state healthcare organisation, e.g. the NHS
...A healthcare related government organisation, e.g. PHE
...A non-healthcare related government organisation
...An independent charity e.g. Cancer Research UK
...A private medical research company

5. Results

5.1 Data Reduction

Responses to completed questionnaires were subjected to Principal Component Analysis with Direct Oblimin Rotation in order to identify the underlying factors and reduce the number of items. Factors with eigenvalues > 1 were extracted, and items with loadings of less than .4 were excluded from the factor.

Consistent with existing literature (see Whiddett et al., 2006), factor analysis identified that participant's responses may be split according to the organisations to which they refer. For example, there may be systematic differences in participants' willingness/happiness to share data depending on whether they are specifically health-related or private companies. However, this finding was not consistent throughout the results, for all measures. For this reason, it was decided that factor loading was not appropriate, and each item was tested separately. Factor loading tables can be seen in Appendix 3. The possible impact of recipient organisation on data sharing is further discussed in the Discussion section.

5.2 Primary Analyses

Analysis was conducted using a repeated measures analysis of variance (ANOVA) to allow an examination of the effects of stigma (high or low), risk/threat (high or low) and data type (anonymous, linked-identifiable or identifiable), as well as interactions between these variables. Within this analysis there were three factors (stigma, risk/threat, and data-type), with two levels of stigma (high vs low), two levels of risk/threat (high vs low) and three levels of data type (anonymous vs. linked-identifiable, vs identifiable). The repeated measures analyses also considered the interaction between these factors. All simple effects analyses (i.e., further analysis of interaction effects and main effects) were conducted using Bonferroni corrections. All data were analysed using IBM SPSS version 25.

As per the structure of the survey, all items coded lower denote greater willingness to share or acceptability of data-sharing. For example, as on a 5-point Likert scale, 1 is equal to very accepting of data-sharing where 5 equals not at all accepting of data sharing. As per convention, interaction and main effect analyses that do not meet significance (i.e., $p \geq .05$) are not discussed further within this manuscript.

Necessary. There was a marginally significant effect of stigma on the perception of data sharing as necessary, $F(1, 56) = 3.49$, $p=0.067$.

There was a significant main effect of data type on necessity of data sharing, $F(1.24, 69.64) = 72.08$, $p<0.001$: Participants saw data sharing as more necessary if the data was anonymised ($M=1.98$) rather than linked-identifiable ($M=2.29$, $p < .001$) or identifiable ($M = 3.23$, $p < .001$), and if the data was linked-identifiable rather than potentially identifiable ($p < .001$).
There was a marginally significant interaction effect between risk/threat and data type $F(1.81, 101.32) = 3.05$, $p=0.057$.

There was a significant interaction effect between stigma and data type $F(2, 112) = 4.28$, $p=0.016$. As reported in the main effects analyses, individuals perceive data sharing as more necessary under conditions of high stigma relative to low stigma, however this effect only holds when the data is anonymous ($p = .006$) and not when the data is linked-identifiable ($p = .307$) or identifiable ($p = .393$). See Figure 1.

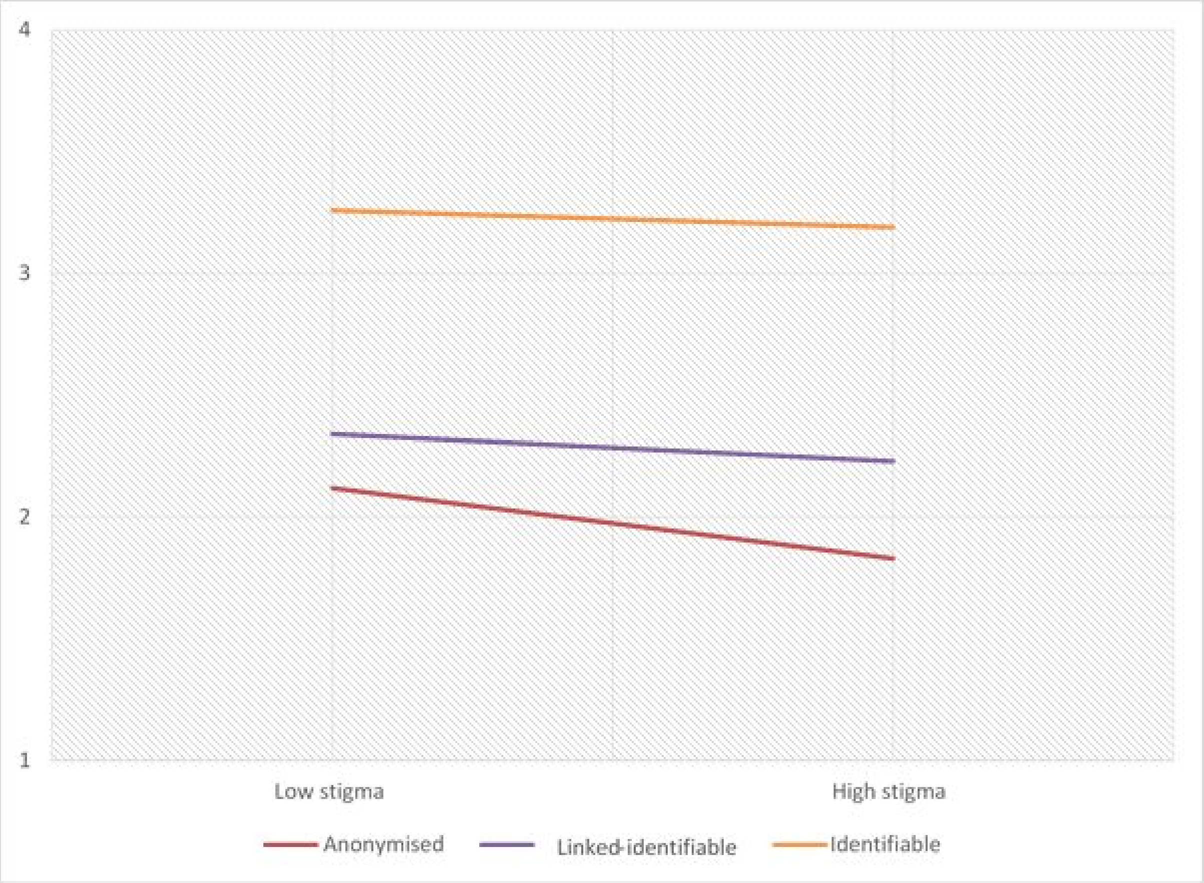

Figure 1. *Interaction between stigma and data type on perceived necessity of data sharing*

There was no main effect of risk/threat and no interactions between risk/threat and stigma or between risk/threat, stigma, and data type, all Fs < 1.456, ps > .233.

Urgency. There was a significant main effect of stigma on the perceived urgency of data sharing F(1, 56) = 15.91, p > .001. Participants perceived data sharing to be more urgent following the high stigma scenarios (M=2.98) compared to the low stigma scenarios (M=3.26, p < .001).

There was a significant main effect of risk/threat on the perceived urgency of data sharing, F(1, 56) = 8.95, p = 0.004: participants perceived data sharing as more urgent within the high risk/threat scenarios (M=2.98) rather than low risk/threat scenarios (M=3.26, p = .004).

There was a significant main effect of data type on the perceived urgency of data sharing, F(1.31, 73.41) = 36.39, p < .001. Participants perceived data sharing as more urgent when it was anonymised (M = 2.83) rather than linked-identifiable (M = 2.96, p = 0.03) or identifiable (M = 3.58, p < .001); and linked-identifiable rather than identifiable (p < .001).

There were no interaction effects between risk/threat, stigma and data type: Fs < 2.11, ps > .152.

'I would be happy for my data to be shared'. There was a significant effect of risk/threat on participants' happiness to share data: F(1, 56) = 5.83, p = 0.019. Participants were happier to share data under high risk/threat circumstances (M = 2.92) compared to low risk/threat (M = 2.45).

There was a significant effect of data type on the happiness of participants for sharing their personal data: F(1.44, 80.79) = 83.69, p < .001. Participants were happier to share when data was anonymised (M = 1.81) rather than linked-identifiable (M = 2.18, p < .001) or identifiable (p < .001), and linked-identifiable rather than identifiable (M = 3.12, p < .001).

There was no main effect of stigma and no interaction effects, all Fs < .119 and ps < .250.

'I would be happy for my data to be shared if my consent were requested from my doctor/other healthcare professional.' There was a significant effect of data type on the happiness of sharing data, when consent was requested from healthcare professionals: F(1.32, 74.16) = 35.56, p < .001. Participants were happier to share their personal data when it was anonymised (M = 1.60) rather than linked-identifiable (M = 1.82, p = 0.001) or identifiable (p < .001), and linked-identifiable rather than identifiable (M = 2.42, p < .001).

There was a marginal significant effect of risk/threat on participants' happiness of sharing data, when consent was requested from healthcare professionals: F(1, 56) = 3.34, p = 0.073.
There was no main effect of stigma and no interaction effects, all Fs < .138 and ps < .237.

'I would be happy for my data to be shared without my explicit consent.' There was a significant effect of stigma on participants' happiness to share data without explicit consent: F(1, 56) = 5.73, p = 0.020. Under conditions of low stigma, participants were happier to share their data (M = 3.16) compared to high stigma (M = 3.37).

There was a significant effect of data type on the sharing of data without explicit consent: F(1.32, 74.16) = 35.56, p < .001. Participants were happier if the data was anonymised (M = 1.60) rather than linked-identifiable (M = 1.82, p = 0.001) or identifiable (p < .001) and linked-identifiable rather than identifiable (M = 2.42, p < .001).

There was no main effect of risk/threat and no interaction effects, all Fs < 0.11 and ps < .403.

'I would be happy for my data to be shared with my doctor.' There was a significant effect of data type on participants' happiness towards sharing with their doctor: F(1.33, 74.45) = 12.97, p < .001. There was no significant difference between sharing data that was anonymised (M = 1.38) and linked-identifiable (M = 1.48, p = 0.094); however, there was a significant difference between sharing data that is anonymised versus identifiable (M = 1.79, p = 0.001), and a significant difference between sharing data that is linked-identifiable versus identifiable (p = 0.004). It is important to note that means for happiness to share with doctor, in each different risk/threat, stigma and data-type scenario, were the lowest for all outcomes (were all below the value of 2.00), denoting that participants, on average, either 'agreed' or 'strongly agreed' to share their data with this group.

There were no main effects of risk/threat and stigma, and no interaction effects, all Fs < .063 and ps < .214.

'I would be happy for my data to be shared with academic researchers from a university.' There was a significant effect of data type on the sharing of data with academic researchers from a university: F(1.32, 74.16) = 35.56, p < .001. Participants were happier if the data was anonymised (M = 1.87) rather than linked-identifiable (M = 2.15, p < .001) or identifiable (p < .001) and linked-identifiable rather than identifiable (M = 2.95, p < .001).

There were no main effects of risk/threat and stigma, and no interaction effects, all Fs < .229 and ps < .162.

'I would be happy for my data to be shared with a state healthcare organisation, e.g. NHS.' There was a significant effect of data type on the sharing of data with a state healthcare organisation: F(1.31, 73.53) = 37.63, p < .001. Participants were happier if the data was anonymous (M = 1.66) rather than linked-identifiable (M = 1.88, p < .001) or identifiable (p < .001) and linked-identifiable rather than identifiable (M = 2.49, p < .001).

There was a significant interaction effect between risk/threat and data type F(2, 112) = 3.08, p = 0.05. For the high-risk/threat scenarios, the previous effect of data type on happiness to share data was observed: participants were happier to share anonymised data (M = 1.61) than linked-identifiable data (M = 1.91, p < .001) or identifiable data (M = 2.47, p < 0.001), and happier to share linked-identifiable data than identifiable data (p < .001). For the low risk/threat scenarios, the expected difference was observed between anonymous data (M = 1.72) and identifiable data (M = 2.52, p < .001), and between linked-identifiable data (M = 1.86) and identifiable data (p < .001). However, there was no significant difference in happiness to share anonymous or linked-identifiable data (p = 0.72). See Figure 2.

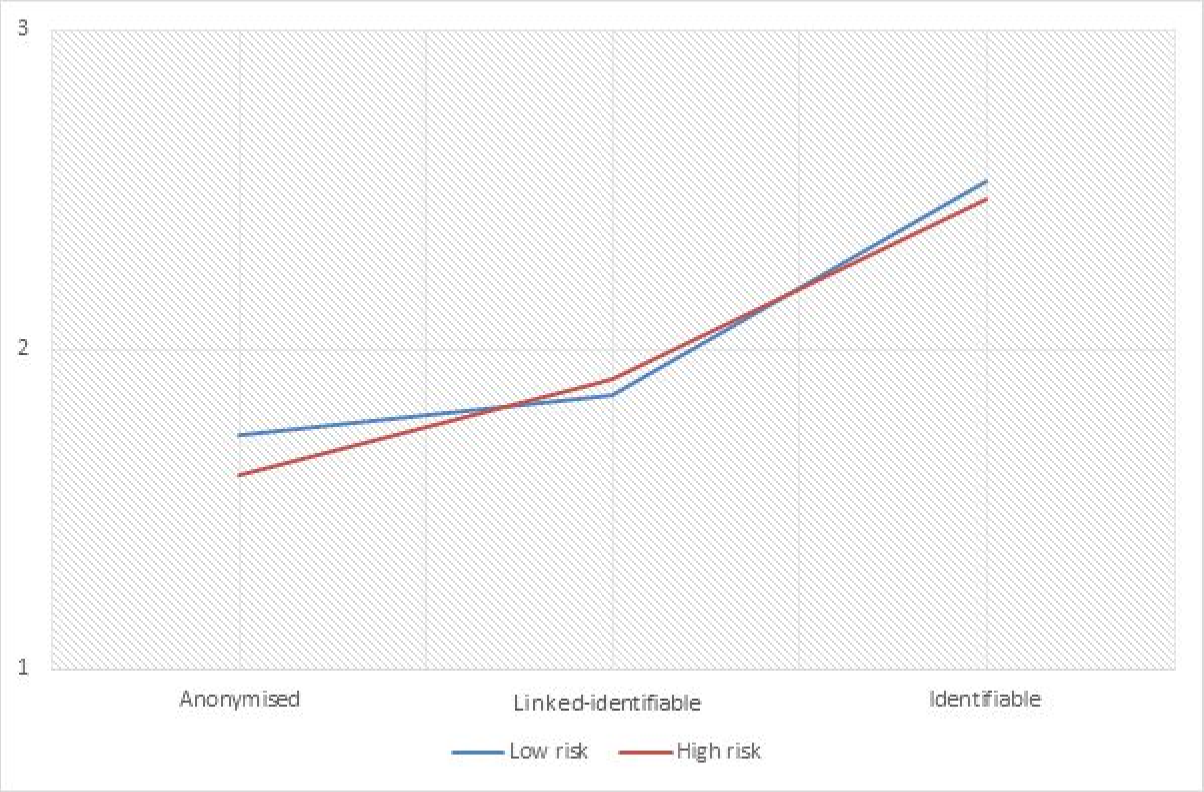

Figure 2. *Interaction between stigma and data type on perceived happiness to share data with a state healthcare organisation, such as the NHS*

There were no effects of stigma or risk/threat and no further interaction effects, all Fs > 0.92, all ps > 0.39.

'I would be happy for my data to be shared with a healthcare related government organisation, e.g. Public Health England.' There was a significant effect of data type on the sharing of data with a healthcare related government organisation: F(1.36, 76.36) = 47.12, p < .001. Participants were happier if the data was anonymised (M = 1.79) rather than linked-identifiable (M = 2.06, p < .001) or identifiable (p < .001) and linked-identifiable rather than identifiable (M = 2.70, p < .001).

There were no main effects of risk/threat and stigma, and no interaction effects, all Fs < .029 and ps < .164.

'I would be happy for my data to be shared with a non-healthcare related government organisation.' There was a significant effect of stigma on participants' happiness to share data with non-health related government organisations: F(1, 56) = 6.65, p = 0.013. Participants were happier to share their data under conditions of low stigma (M = 2.92) compared to high stigma (M = 3.07).

There was a significant effect of data type on the sharing of data for this measure: F(1.51, 84.47) = 54.49, p < .001. Participants were happier if the data was anonymised (M = 2.62) rather than linked-identifiable (M = 2.84, p < 0.003) or identifiable (p < .001) and linked-identifiable linked-identifiable rather than identifiable (M = 3.52, p < .001).

Finally, there was a marginally significant, three-way interaction between risk/threat, stigma and data, F(1.18, 22.66) = 2.91, p = 0.058.

There was no main effect of risk/threat, and no further interaction effects, all Fs < .012 and ps < .320.

'I would be happy for my data to be shared with an independent charity, e.g. Cancer Research UK.' There was a significant effect of data type of participants' happiness to share with an independent charity, such as Cancer Research: F(1.48, 81.37) = 47.56. Participants were happier to share anonymised data (M = 2.48) versus linked-identifiable (M = 2.75, p = 0.001) or identifiable (M = 3.46, p < .001), and linked-identifiable rather than identifiable (p < .001).

There was a significant interaction effect between stigma and data, F(2,110) = 3.19, p = 0.045, such that individuals were happier to share data under conditions of low stigma relative to high stigma, but only when the data was identifiable (3.39 vs. 3.55, p = 0.012); no difference was observed for either anonymous (2.51 vs 2.46, p = 0.56) or linked-identifiable identifiable (2.76 vs 2.76, p = 1.00) data. See Figure 3.

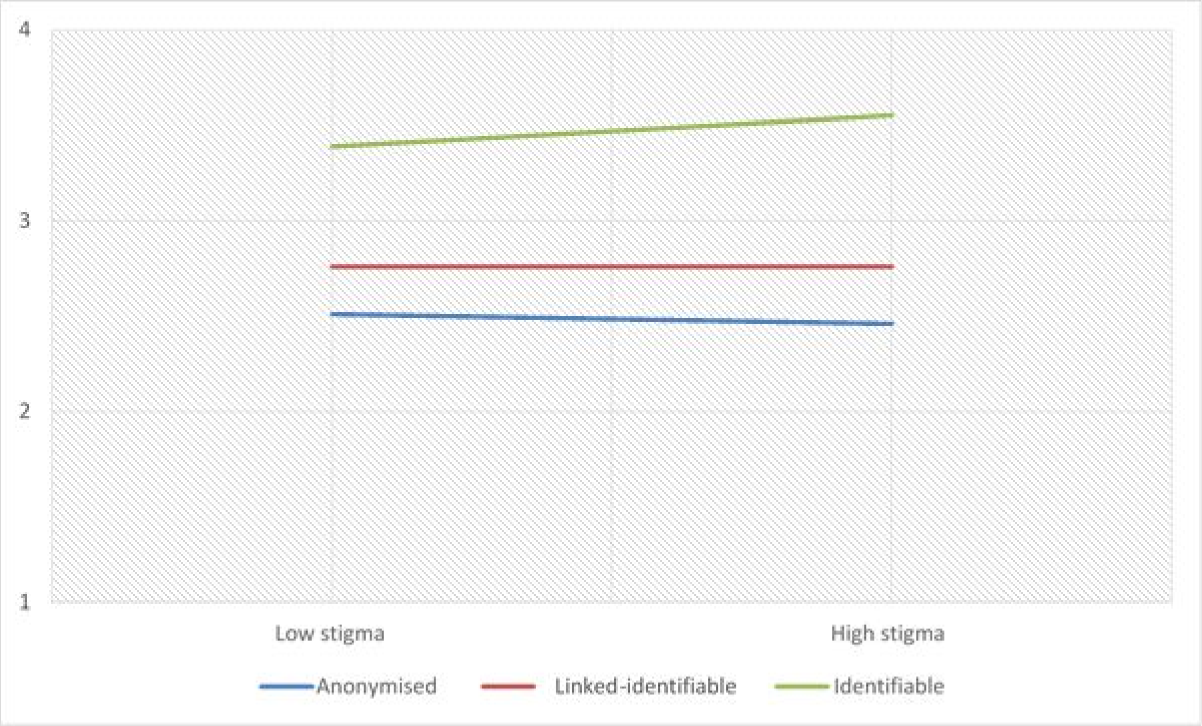

Figure 3. *Interaction between stigma and data type on perceived happiness to share data with an independent charity*

There were no effects of stigma or risk/threat and no further interaction effects, all Fs > 0.003, all ps > .271.

'I would be happy for my data to be shared with a private medical research company.' There was a significant effect of stigma on the happiness of sharing with private medical companies: F(1.31, 70.91) = 34.67. Participants were happier to share data that was associated with low stigma (M = 3.37) than high stigma (M = 3.52, p = 0.003).

There was a significant effect of data on participants' happiness to share data with private medical research companies: F(1.31, 70.91) = 34.67. Participants were happier to share data that was anonymous (M = 3.13) rather than linked-identifiable (M = 3.33, p = 0.005) or identifiable (M = 3.88, p < .001), and happier to share linked-identifiable data rather than identifiable (p < .001). It is important to note that this measure had the highest means, inferring that participants were most unhappy about sharing with these kinds of organisations. This is discussed further in the next section.

There was no effect of risk/threat and no further interaction effects, all Fs > .114, all ps > .148.

5.3 Secondary Analyses

Following the completion of the primary analyses and discussions with the WP 2 lead, the conduct of further analysis was discussed to provide some insight into the impact of data recipient on participants' happiness to share data. This provided the opportunity to further explore the impact of data recipient on willingness as discussed in the introduction(e.g., Barrett et al., 2006; Whiddett et al., 2006).

This further supplementary analysis has involved averaging across participants responses to items concerning their happiness to share anonymous, linked-identifiable and identifiable data with each of the different data recipients across each scenario. This resulted in seven scores, one for each data recipient (Doctor, University, state healthcare organisation, healthcare related government organisation, non-healthcare related government organisation, independent charity, private medical research company), which were subjected to a single factor ANOVA analysis to examine differences in happiness. As before, all simple effects analyses were conducted using Bonferroni corrections; all analysis was conducted using IBM SPSS v25. Means and standard deviations (S.D.) are in Table 3 (as before, lower scores = greater happiness).

Table 3. *Means and standard deviations for all aggregate responses*

I would be happy for my data to be shared with the following organisations...	Mean	S.D.
...My Doctor	1.56	0.72
...Academic researchers from a University	2.33	0.88
...A state healthcare organisation, e.g., the NHS	2.02	0.89
...A healthcare related government organisation, e.g. PHE	2.19	0.89
A non-healthcare related government organisation	2.99	0.99

...An independent charity, e.g., Cancer Research UK	2.92	1.01
...A private medical research company	3.49	1.11

Analysis revealed a significant main effect of data recipient on happiness to share data, $F(3.548, 198.674) = 71.51$, $p < .001$. Follow up contrast analysis revealed that participants reported significant differences in their happiness to share data with all individual organisations (all ps <.001), except for between universities and healthcare related government organisations ($p = .859$) and between non-healthcare related government organisations and independent charities ($p = 1.00$). Descriptively, and based on the values reported in Table 3, this means that individuals were happiest to share information with these organisations in the following order (from most to least happy):

1. Doctor
2. State healthcare organisation
3. Healthcare related government organisation / universities (no difference between the two)
4. Independent charity/ non-healthcare related government organisation (no difference between the two)
5. Private medical research company

6. Discussion

The present study sought to address the lacunae in research from the patients' perspective (Whiddett et al., 2006) by exploring the impact of risk/threat, stigma, and data type on patients' attitudes towards data sharing. Overall, the results from this study indicate that although there are variable effects of risk/threat and stigma on perceived necessity, urgency and happiness of participants to share information, data type emerged as a consistent predictor. Furthermore, there was a recurring interaction between stigma and data type on both the perceived *necessity of data sharing*, and *happiness to share data with an independent charity*; there was in interaction between risk/threat and data type when sharing with a state healthcare organisation. However, these interaction effects are occasional and inconsistent, so cannot be taken to fully support the hypotheses. In the sections that follow, the results are discussed in the context of the extant literature outlined in the introduction.

6.1 Data type

There was a clear effect of data type on the perceived necessity, urgency and happiness of sharing data: participants were least willing to share identifiable personal information, more willing to share linked-identifiable data and most willing to share when data was fully anonymous. These results reflect the notion that individuals do not wish to be identified from their personal data, particularly when it takes a more personal nature.

Here, maintaining privacy and confidentiality are specific motives for individuals (Whiddett et al., 2006). Whilst pseudonymisation upholds anonymity, the prevailing (albeit small) risk of being identified means individuals wish to share this type of data less than when it is fully anonymous (Ewerlöf, 2018). Although individuals may recognise the functionality of being identified from their health data (e.g., for use in screen and treat programmes or within wider medical research), they are worried about the accidental exposure of sensitive information or the thought of being insidiously surveilled (McGrath, 2004; MRC, 2007; Taylor, 2010; Wellcome Trust, 2015; Dogan, 2015; Jones & Gagneja, 2016). Furthermore, having control of personal information may be prompted by concerns over the consequent misuse of personal data; misuse contributes to the experience of mistrust and a reluctance to share information at all (Kaufman et al., 2009; van Panhuis et al., 2014; Rubin et al., 2018). These sentiments may have been exacerbated by the Cambridge Analytica scandal (in early 2018, earlier in the same year as data collection for this survey began). The high-profile nature of this scandal, and ongoing criticisms of data misuse, may have had a negative impact on questionnaire responses and exacerbate opinions of mistrust and an unwillingness to share (Kaufman et al., 2009; van Panhuis et al., 2014; Granville, 2018; Landau, 2018). Accordingly, anonymous information is most willingly shared and identifiable least; our results provide overwhelming support for these existing findings.

6.2 The perception of risk/threat

To our knowledge, this study is the first to manipulate risk/threat – high versus low – when studying the public acceptability of sharing personal information. As mentioned in the introduction, the public perceive data-sharing to be particularly important after major events wherein this process is deemed as for the ‘greater good’ (Dogan, 2015; Rubin et al., 2018). It was therefore unsurprising that there was a main effect of risk/threat on attitudes towards data-sharing. Although the experience of heightened risk is thought to influence behavioural avoidance (Gigerenzer, 2006), under specific conditions of public health and security emergencies (as presented herein), the reverse appears to be true. Further research is therefore needed to further unpick the specific conditions under which risk/threat may lead to greater or lesser willingness to share data.

For perceived happiness of sharing data with a state healthcare organisation, such as the NHS, there was an interaction effect between risk/threat and data type (influenced by the main effect of data type). In high-risk/threat scenarios individuals were happier to share anonymised data, followed by linked-identifiable, and least happy about sharing identifiable data. In low-risk/threat scenarios, differences were observed between anonymous and identifiable data, and linked-identifiable and identifiable data; however, there was no significant difference between sharing anonymous and linked-identifiable data, such that individuals are more willing to share these types of data when the associated stigma is low. This finding does not support the hypothesis whereby it was thought that the perception of high risk/threat would mitigate the associated worries of being identified from personal data. A previous study found that individuals are willing to share their data within high risk scenarios due to the perceived benefits of being identified from PID after being involved in a major incident (Rubin et al., 2018). However, the present findings demonstrate that individuals are less happy to share identifiable data. Since the manipulation of risk/threat was a more novel element of the present study there is little existing literature on the effect of this. Extending research of this kind may elucidate the full impact of risk/threat.

6.3 The perception of stigma

As hypothesised, individuals are less happy to share highly stigmatised information, but only under the conditions where data is being shared either a) without consent or b) with organisations that are not typical healthcare organisations (i.e., not the NHS or personal GP but other non-healthcare government and private healthcare providers). These findings highlight that there may be an impact of data-recipient particularly when they are not archetypally health-related. In other words, individuals are not concerned about whether the data is stigmatised if it is being shared among their healthcare providers and with their consent; when it is not, members of the public are more likely to be protective of data that is stigmatised (and thus more sensitive).

Within this study, there are two seemingly contradictory interactions between stigma and data type. First, there is an interaction between stigma and data type for happiness to share data with charities – the effect of stigma is only found for identifiable data, being that individuals do not want to share. However, the opposite effect is observed for the perceived necessity of data sharing – participants saw data sharing as more necessary if the data was anonymised. The former effect makes sense in relation to the previous literature. For example, Kaufman and colleagues (2009) found that certain types of medical information require extra privacy, whereby HIV infection, sexual history and mental illness require the most, because of the associated sensitivity.

Perhaps due to the perceived lack of control over such data once shared, individuals share more sensitive personal information with increasing reluctance. Indeed, ensuring anonymity of personal information, particularly that which may be stigmatised, is important for members of the public (Barrett et al., 2006); Willison and colleagues (2009) assert that individuals are reluctant to share stigmatised health data out of concern that it could be used to discriminate against them. Accordingly, in a study of the factors associated with reporting mental health symptoms, when completing questionnaires anonymously (compared to when being identified responses), military personnel were more likely to report more honestly, indicating that barriers to mental health care were it being ‘too embarrassing’ as well as associating them with weakness (Fear et al., 2012).

However, this is inconsistent with the second interaction effect. Given that the former relates to *happiness* to share data and the latter relates to the *necessity* of data sharing, it is possible that these effects relate to the precise questions asked. As such, individuals may recognise the necessity, or ‘public good’, of sharing data (particularly anonymous data, as per existing literature). This is true for stigmatised issues relative to a non-stigmatised, but when it comes to actually sharing their own data they are equally happy to share both unless it is identifiable. Regardless, further work is therefore needed to unpick the precise relationship between stigma and data type and its effect on attitudes towards data sharing.

6.4 Highlighting the importance of context and data-recipient

The previous finding illuminates that there is an impact of who data-recipient is, where previous studies have found that there are differences between healthcare providers and companies which are private (Whiddett et al., 2006). Initial secondary analyses (un-hypothesised) were conducted on the current dataset to examine this possibility, revealing that participants are, broadly, happiest to share personal information with their doctor and least happy to share with private healthcare providers. This finding is consistent with Whiddett and colleagues (2006), who reported that participants express a particular reluctance to share personal information if: a) it is identifiable, and; b) the data-recipient is less involved with the provision of healthcare. This finding is consistent with other literature, insofar as health-related organisations are perceived to be wardens of personal and sensitive information (European Commission Research Group, 2003). Indeed, individuals are concerned about personal information “falling into the wrong hands” (MRC, 2007, p. 8).

Further, more complex analysis to examine the relationship between different types of data (e.g., identifiable etc) and happiness to share data with these organisations (as per Whiddett and colleagues, 2006) was beyond the scope, rationale, hypotheses, and design of the current project, and, as such, purpose built research projects should be conducted in future to explore this possibility.

7. Limitations and future recommendations

An important consideration to make is whether respondents had personal experience of issues raised in the present study. The personal experience of either an extreme event or of mental illness may influence an individual’s happiness towards sharing personal information. For example, the stigma associated with mental illness may contribute to a reluctance to share, but on the other hand may result in an increased willingness, driven by the notion of raising awareness. Alternatively, individuals may have not completed the survey based on this and these responses would have been discounted altogether.

The self-selective, convenience sample used in this report may mean the sample is not wholly representative of the population. As well as having more direct access to the study, it is conceivable that having access to the internet means individuals are generally more accepting of sharing personal information (Rubin et al., 2018). The use of a recruitment company for future work may help to mitigate some of these concerns.

Increasing awareness of potential data-uses and the associated vocabulary, may bolster willingness to share personal data. Being poorly informed about these factors may reduce willingness to share (Willison et al., 2009; Whiddett et al., 2006; Riordan et al., 2015). For example, the public may not know exactly what the terms ‘*data-linking*’, *‘linked-identifiable’* or ‘*pseudonymised’* refer to. Indeed, Hull and colleagues (2008) identify that the difference between identifiable and anonymised information may not be meaningful to patients, and Whiddett et al. (2006) state that there may be public confusion as to why and what level of consent occurs. Future research may be required to understand the public's current perceptions of anonymisation, consent and how personal information can be integrated and used, and if these alter the public acceptability of sharing data.

A series of pilot tests identified alcohol dependency to be particularly stigmatised amongst the public (see Appendix 1) and was integrated into vignettes because it elicited this specific response. However, it may be interesting to explore whether other but stigmatised conditions would elicit the same response. For example, alcohol dependence is less commonly regarded as a mental illness compared to depression or schizophrenia; this is due to its perceived 'controllability' wherein those who experience alcohol dependence are responsible for their illness (Schomerus et al., 2010). As such, public attitudes are believed to be illness-specific (Link et al., 1999; Angermeyer & Matschinger, 2003). Future research could therefore assess the public acceptability and perceptions of sharing data related to a spectrum of conditions and illnesses.

8. Conclusion

These results indicate that data type has a definite impact on the happiness of individuals to share their personal information. Individuals are increasingly reluctant to share their personal information when it becomes more sensitive, unless they have guaranteed anonymity. This corresponds to the existing literature (e.g. Barrett et al., 2006; Whiddett et al., 2006). The development and strengthening of the GDPR may mitigate many of these problems for individuals, but this needs to be explored further.

The main effect of risk/threat on attitudes towards data-sharing, present within the study, highlights that the public are happier to share their data after major events (Dogan, 2015; Rubin et al., 2018). Additionally, individuals are less happy to share highly stigmatised information, in particular when sharing with organisations that are not related to traditional forms of delivering healthcare. Where these results were inconsistent, more information should be gathered to understand the true impact of risk/threat and stigma. For example, stigma had variable effect on both the perceived happiness and necessity of sharing personal information, but this outcome could be as a result of the particular wording in questionnaires.

Whilst personal information exists as an important resource and may be instrumental in innovating how healthcare is provided, individuals are reluctant to share their data out of fear that it will be erroneously used (Kaufman et al., 2009; Willison et al., 2009). Preliminary additional analysis reveals that there are additional differences in perceptions dependent upon the data recipient (i.e., healthcare provider vs. private company), which should be scrutinised further in future research.

9. References

Abdallah, S., & Panjabi, R. (2008). Chapter 6: Epidemiology and Surveillance. In G. Burnham, & E. Rand, (Eds.), The Johns Hopkins and Red Cross Red Crescent Public Health Guide in Emergencies, Second Edition (pp. 220-284), Johns Hopkins School of Hygiene and Public Health, Baltimore, MD.

Abouzahr, C., Adjel, S., & Kanchanachitra, C. (2007). From data to policy: good practices and cautionary tales. *Lancet, 369*(9566), 1039-1046.

Angemeyer, M. C. & Matschinger, H. (2003). The stigma of mental illness: effects of labelling on public attitudes towards people with mental disorder. *Acta Psychiatr Scand, 108*(4), 304-9.

Astrup, J. (2018). GDPR: The transfer of data power. *Community Practitioner, 91*(5), 36-39. Retrieved from https://search.proquest.com/docview/2056424236?accountid=11862

Barrett, G., Cassell, J. A., Peacock, J. L., & Coleman, M. P. (2006). National survey of British Public's views on use of identifiable medical data by the National Cancer Registry. *BMJ, 332*(7549), 1068-1072.

Beskow, L. M., & Dean, E. (2008). Informed consent for biorepositories: assessing prospective participants' understanding and opinions. *Cancer Epidemiol. Biomarkers Prev, 6*, 1440-1451.

Boerma, J. T., & Stansfield, S. K. (2007). Health statistics now: are we making the right investments? *Lancet, 369*(9563), 779-786.

Brown, V. J. (2014) Risk Perception: It's personal. *Environmental Health Perspectives, 122*(10). A276-A279.

Buchwald, D., Mendoza-Jenkins, V., Croy, C., McGough, H., Bezdek, M., & Spicer, P. (2006). Attitudes of urban American Indians and Alaska Natives regarding participation in research. *Journal of General Internal Medicine, 21*(6), 648-651.

Chan, M., Kazatchkine, M., Lob-Levyt, J., Obaid, T., Schweizer, J., Sidibe, M., Veneman, A., & Yamada, T. (2010). Meeting the demand for results and accountability: a call for action on health data from eight global health agencies. *Plos Med, 7*(1), e2000223.

Chew, C. & Eysenbach, G. (2010) Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. *PLOS ONE 5*(11): e14118.https://doi.org/10.1371/journal.pone.0014118

Clickbait (2019). In *English Oxford Living Dictionaries.* Retrieved from https://en.oxforddictionaries.com/definition/clickbait

Close, R. M., Maguire, H., Etherington, G., Brewin, C. R., Fong, K., Saliba, V., Barker, R. M., & Leonardi, G.S. (2014). Preparedness for a major incident: Creation of an epidemiology protocol for a health protection register in England. *Environment International, 72,* 75-82.

Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist, 59*(7), 614-625.

Damschroder, L. J., Pritts, J. L., Neblo, M. A. Kalarickal, R. J., Creswell, J. W., & Hayward, R. A. (2007). Patients, privacy, and trust: patients' willingness to allow researchers to access their medical records. *Soc. Sci. Med., 1*, 223-235.

Data Protection Act 1998 (1998), London: Stationery Office.

Dogan, A. I. (2015). *A Qualitative Study Investigating Public Attitudes Towards Health Protection Registers Established after a Major Incident* (Masters Thesis). Department of Psychological Medicine. London, King's College London.

Dye, C., Bartolomeos, K., Moorthy, V, & Kieny, M. P. (2016). Data sharing in public health emergencies: a call to researchers. *Bull World Health Organ. 94*, 158.

European Commission Research Group (2003). *Data protection. European Commission Research Group.* Brussels: Directorate General Press and Communication.

Ewerlöf, A. (2018, May 31). GDPR pseudonymisation techniques. *Medium.* Retrieved from https://medium.com/@alexewerlof/gdpr-pseudonymization-techniques-62f7b3b46a56

Fear, N. T., Seddon, R., Jones, N., Greenberg, N., & Wessely, S. (2012). Does anonymity increase the reporting of mental health symptoms? *BMC Public Health. 12*(797), 1-7.

Fife, B. L. & Wright, E. R. (2000). The demensionality of stigma: a comparison of its impact on the self of persons with HIIV/AIDS and cancer*. J. Health Soc. Behav. 41*, 50-67.

Gigerenzer, G. (2006). Out of the Frying Pan into the Fire: Behavioural Reactions to Terrorist Attacks. *Risk Analysis. 26*(2), 347-351.

Gobin, M., Rubin, G. J., Albert, I., Beck, A., Danese, A., Greenberg, N., … Oliver, I. (2018). Outcomes of Mental Health Screening for United Kingdom Nationals Affected by the 2015-2016 Terrorist Attacks in Tunisia, Paris, and Brussels. *Journal of Traumatic Stress. 00*, 1-9.

Granville, K. (2018, March 19). Facebook and Cambridge Analytica: What you need to know as fallout widens. *The New York Times.* Retrieved from https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html

Hill, E. M., Turner, E. L., Martin, R. M., & Donovan, J. L. (2013). “Let’s get the best quality research we can”: public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. *BMC Medical Research Methodology. 13*(72), 1-10.

HM Government (2007). *Data protection and sharing - guidance for emergency planners and responders.* London: Cabinet Office.

Hull, S. C., Sharp, J. R., Brown, M., Hughes, M., Sugarman, J., Schwinn, D., … Katz, T. (2008). Patients' Views on Identifiability of Samples and Informed Consent for Genetic Research. *Am J Bioeth. 8*(10), 62-70.

Information Commissioner's Office (2018). *Guide to the General Data Protection Regulation (GDPR).* Retrieved from https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf

Ipsos, M. (2014). *Public attitudes to the use and sharing of their data. Research for the Royal Statistical Society by Ipsos MORI.* London: Ipsos MORI.

Jones, A.S. & Gagneja, K. (2016). Preventing covert webcam hacking in civilian and governmental sectors. *International Conference on Computational Science and Computational Intelligence (CSCI)*, 993-998. doi: 10.1109/CSCI.2016.189

Kaufman, D. J., Murphy-Bollinger, J., Scott, J., & Hudson, K. L. (2009). Public opinion about the importance of privacy in biobank research. *The American Journal of Human Genetics, 85*, 643-654.

Keahey, L. (2018). *"Millennial" Attitudes Towards Sharing Mobile Phone Location Data with Health Agencies: A Qualitative Study. (Masters thesis).* King's College London, London.

Klingemann, H. (2001). *Alcohol and its social consequences: the forgotten dimension.* Copenhagen: WHO Regional Office for Europe.

Kostkova, P. (2018). Disease surveillance data sharing for public health: the next ethical frontiers. *Life Sciences, Society and Policy. 14(*16), 1-5.

Landau, S. (2018). Programming programming languages, and analysing Facebook's failure. *Communications of the ACM, 61*(6), 8-9.

Link, B.G. & Phelan, J.C. (2001). Conceptualising Stigma. *Annual Review of Sociology. 27*, 363-385.

Link, B.G. & Phelan, J. C. (2006). Stigma and its public health implications. *Lancet. 367*(9509), 528-9.

Logie, C. & Gadalla, T. M. (2009). Meta-analysis of health and demographic correlates of stigma towards people living with HIV. *AIDs Care. 21*(6), 742-753.

McGrath, J. E. (2004). *Loving big brother: Performance, privacy, and surveillance space.* London: Routledge.

Medical Research Council (2007). *The Use of Personal Health Information in Medical Research.* Retrieved from https://mrc.ukri.org/documents/pdf/the-use-of-personal-health-information-in-medical-research-june-2007/

Murdoch, T. B. & Detsky, A. S. (2013). The Inevitable Application of Big Data to Health Care. *JAMA. 309*(13), 1351-1352.

Murray, C. J. (2007). Towards good practice for health statistics: lessons from the Millennium Development Goal health indicators. *Lancet, 369*(9564), 862-873.

O'Donnell, Corrigan and Gallagher (2015). The impact of anticipated stigma on psychological and physical health problems in the unemployed group*. Frontiers in Psychology. 6*, 1263.

Page, S. A., & Mitchell, I. (2006). Patients' opinions on privacy, consent and the disclosure of health information for medical research. *Chronic Dis. Can., 2*, 60-67.

Peters, E.M., Burraston, B. & Mertz, C. K. (2004). An emotion-based model of risk perception and stigma susceptibility: cognitive appraisals of emotion, affective reactivity, worldviews, and risk perceptions in the generation of technological stigma*. Risk Analysis. 24*(5), 1349-1367.

Peto, J., Fletcher, O., & Gilham, C. (2004). Data protection, informed consent, and research*. BMJ, 328,* 1029-30.

Phelan, J. C., Link, B. G., Stueve, A. & Pescosolido, B. A. (2000). Public Conceptions of Mental Illness in 1950-1996: What is mental illness and is it to be feared? *Journal of Health and Social Behaviour. 41*(2), 188-207.

Poikola, A., Kuikkaniemi, K., & Honko, H. (2014). *MyData – A Nordic Model for human-centred personal data management and processing.* [White paper]. Retrieved from https://www.lvm.fi/documents/20181/859937/MyData-nordic-model/2e9b4eb0-68d7-463b-9460-821493449a63?version=1.0

Riordan, F., Papoutsi, C., Reed, J. E., Martson, C., Bell, D., & Majeed, A. (2015). Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK. *International Journal of Medical Informatics. 84*, 237-247.

Robling, M. R., Hood, K., Houston, H., Pill, R., Fay, J., & Evans, H. M. (2004). Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study. *J. Med. Ethics, 1*, 104-109.

Rubin G.J., Webster R., Rubin A.N., Amlot, R., Grey, N., & Greenberg, N. (2018). Public attitudes in England towards the sharing of personal data following a mass casualty incident: A cross-sectional study. *BMJ Open. 8*, e022852. doi: 10.1136/bmjopen-2018-022852

Schomerus, G., Lucht, M., Holzinger, A., Matschinger, H., Carta, M. G. & Angermeyer, M. C. (2010). The Stigma of Alcohol Dependence Compared with Other Mental Disorders: A Review of Population Studies. *Alcohol and Alcoholism, 46*(2), 105–112.

Siegrist, M., Kelller, C., Kastenholz, H., Frey, S. & Wiek, A. (2007). Laypeople’s and Expert’s Perception of Nanotechnology Hazards. *Risk Analysis. 27*(1), 59-69.

Singer, E., Mathiowetz, N.A. & Couper, M.P. (1993). The Impact of Privacy and Confidentiality Concerns on Survey Participation the Case of the 1990 Us Census. *Public opinion quarterly, 57*(4), 465-482.

Slovic, P. (1987). Perception of Risk. *Science. 236*, 280-285.
Taylor, E. (2010). I spy with my little eye: The use of CCTV in schools and the impact on privacy. *The Sociological Review, 58*(3), 381-405. doi: 10.1111/j.1467-954X.2010.01930.x

Thacker, S. B., Qualters, J. R., & Lee, L. M. (2012). Public health surveillance in the United States: evolution and challenges. *Morb Mortal Wkly Rep Surveil Summ, 61*(suppl), 3-9.

Varotto, S. (2015). The European general data protection regulation and its potential impact on businesses: Some critical notes on the strengthened regime of accountability and the new sanctions. *Communications Law, 20*(3), 78-85. Retrieved from https://search.proquest.com/docview/1720974483?accountid=11862

van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., Heymann, D., & Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health, 14*(1), 1144.

Wallhed Finn, S., Bakshi, A.S. & Andréasson, S. (2014). Alcohol consumption, dependence, and treatment barriers: perceptions among nontreatment seekers with alcohol dependence. *Subst Use Misuse. 49*(6):762-9. doi: 10.3109/10826084.2014.891616.

Wellcome Trust (2015). *The One-Way Mirror: Public attitudes to commercial access to health data.* IPSOS Mori: London.

Whiddett, R., Hunter, I., Engelbrecht, J., & Handy, J. (2006). Patients' attitudes towards sharing their health information. *Int. J. Med. Inform., 7*, 530-541.

Willison, D. J., Steeves, V., Charles, C., Schwartz, L., Ranford, J., Agarwal, G. … Thabane, L. (2009). Consent for use of personal information for health research: Do people with potentially stigmatizing health conditions and the general public differ in their opinions? *BMC Medical Ethics. 10*(10), 1-12.

Wong, M. L., Chia, K. S., Yam, W. M., Teodoro, G. R., & Lau, K. W. (2004). Willingness to donate blood samples for genetic research: a survey from a community in Singapore. *Clin. Genet., 1,* 45-51.

Wong, J. C. (2018, September 29). Facebook says nearly 50m users comprised in huge security breach. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2018/sep/28/facebook-50-million-user-accounts-security-berach

Zakharov, A. (2018, November 2). Private messages from 81,000 hacked Facebook accounts for sale. *BBC News.* Retrieved from https://www.bbc.co.uk/news/technology-46065796

10. Appendices

10.1 Appendix 1

Pilot testing outcomes. To use as justification for choosing alcohol dependence.

Participants & Design
18 employees at Public Health England each completed an online questionnaire in which they were asked to read eight different health and security related scenarios, and then subsequently| to rate these for stigma and dread risk.

Responses to each scenario were statistically compared to help identify and construct scenarios for the primary survey that related clearly to high and low threat/ risk, and to high and low stigma.

Materials & Procedure
All participants read the following eight scenarios, designed to elicit differing levels of stigma and dread risk:

Table 1. *Scenarios presented to pilot participants*

Context	Scenario Text
Depression	Depression is the second leading cause of disability worldwide and is one of the most common mental health disorders in the UK; 10% of the population will experience clinical depression at some point in their lives. Imagine that you have experienced some symptoms associated with mild to moderate depression. You visited your GP about this and they referred you to a specialist support team and you are now recovering well.
Fire	In 2016/17, fire services attended over 30,000 fires at domestic dwellings in England alone. 261 fatalities were reported, predominantly due to either burns, becoming overcome by gas or smoke, or a combination of the two. Imagine that there has been an accidental fire at your house and the Emergency Services have attended. You were not badly injured, but you were taken to the hospital to be treated

	for smoke inhalation. You were discharged on the same day and there were no lasting effects.
Fracture	Fractures are a common ailment in the UK, with approximately 4% of the population presenting with a broken bone every year. Imagine that you have fallen down the stairs and have fractured your ankle. Fortunately, it is a clean fracture and no surgery is required. You ankle is in a plaster cast and you are well on the way to recovery.
Chemical incident	A chemical incident occurs when there is an uncontrolled release of a toxic substance, potentially causing harm to the public and/or the environment. Imagine that you work at a factory that produces potentially hazardous chemicals. There has been a fire at the factory and you are among several people who have been injured. You were not seriously injured, but you did sustain some relatively minor burns for which you were taken to the hospital. Your burns were cleaned and dressed and are now healing well.
Cancer	Cancer is the leading cause of death in Under 75s in the UK. Imagine that you have been diagnosed with a form of skin cancer. Fortunately, it was discovered in the early stages and treatment has been successful. You have been told that you are now in remission.
Terrorist Attack	Following a major incident during which a person may have been exposed to risk of injury/death, or have witnessed the injury/death of other people (e.g., a terrorist attack), it is normal for them to experience some mental distress. When provided with appropriate care and support, the majority of people do not go on to experience significant long term issues.

	Imagine that you have witnessed a terrorist attack, during which a lone gunman attacks a crowd of people. You have not been physically hurt, but you witnessed the injury of several other people. As a result of this you experienced some symptoms of mental distress and visited your GP as a result of this. Your GP referred you to a specialist support team and you are now recovering well.
Airplane Crash	Air travel is, statistically, one of the safest ways to travel. The odds of an aircraft crashing are around 1 in 1.2 million, whilst the odds of dying in such a scenario are around 1 in 11 million. For context, the odds of dying as a result of a road traffic accident are 1 in 5000. Imagine that you were involved in a plane crash. Fortunately, very few people were seriously injured and you were not physically hurt at all. However, since this has happened you have been experiencing regular panic attacks, which you have visited your GP about. Your GP referred you to a specialist support team and you are recovering well.
Alcohol Dependence	Alcohol dependence is prevalent in the UK; 1.9% of adults are considered to be harmful or mildly dependent drinkers, with a further 1.2% meeting the criteria for alcohol dependence. When provided with appropriate care and support, many people show an improvement. Imagine that you are concerned about the amount of alcohol that you are drinking and the effects that this is having on your life. You visited your GP about this and they referred you to a specialist support team. You are now recovering well.

Following the presentation of each scenario, participants were presented with eight statements drawn from previous literature, corresponding to stigma and dread risk, and were asked to indicate their degree of agreement with each (1 = strongly agree; 5 = strongly disagree). All items are presented in Table 2.

Table 2. *Questionnaire items used to represent dread risk and stigma.*

Construct	Items
Dread Risk	This is a risk that harms one person at a time
Dread Risk	This is a risk that could cause harm to a large number of people at once (*)
Dread Risk	This is a risk that most people have learned to live with and would be able to think about reasonably calmly
Dread Risk	Most people would think about this scenario with great dread (*)
Stigma	I would worry about telling people if this happened to me (*)
Stigma	People would be understanding if this happened to me
Stigma	I would be discriminated against if this happened to me (*)
Stigma	I would not be embarrassed if this had happened to me

All four items relating to stigma and all four items related to dread risk were averaged to create single values for each construct. Prior to aggregation, the items marked with an (*) were reverse coded to ensure that positive scores correspond to both greater stigma and greater dread risk. on each value were stigma stigma items were collapsed subsequently coded such that higher values.

Results & Conclusions

Within subjects ANOVA analyses were conducted to compare participants responses to: a) stigma, and; b) dread risk questionnaire items. Specifically, pairwise comparisons were conducted to examine the difference in levels of stigma and dread risk across the scenarios.

Stigma. Mean values and standard errors for participants perceived stigma in response to each scenario are presented in Table 3. Descriptively, the scenario that elicited the highest stigma was the alcohol dependence scenario, while the fracture scenario elicited the lowest stigma.

Table 3. *Means and standard errors for perceived stigma in response to each scenario.*

Scenario	Mean	Standard Error
Depression	3.18	0.17
Fire	2.03	0.13
Fracture	1.99	0.14
Chemical Incident	2.09	0.14
Cancer	2.29	0.16
Terrorist Attack	2.25	0.14
Airplane Crash	2.36	0.19
Alcohol Dependence	4.09	0.18

Inferential analysis revealed a significant difference in perceived stigma across scenarios, $F(7, 112) = 35.18$, $p < .001$. In order to identify whether any scenarios were, statistically (rather than descriptively), indistinguishable in perceived stigma when compared to the descriptively high and low stigma condition (i.e., where there are descriptive differences in mean values which are not significant), we conducted pairwise comparisons between all scenarios and: a) the high stigma scenario (alcohol dependence), and; b) the low stigma scenario (fracture). The depression ($p < .001$), cancer ($p = .035$), airplane crash ($p = .043$) and alcoholism ($p < .001$) all elicited significantly higher perceived stigma than the low stigma scenario (all other *ps* > .14). All scenarios elicited lower stigma than the alcohol dependence scenario (all *ps* ≤ .001). In short, the alcohol dependence scenario was therefore the most stigmatised, with the fracture, terrorist attack, fire, and chemical incident scenarios as jointly the least stigmatised.

Dread risk. Mean values and standard errors for participants perceived dread risk in response to each scenario are presented in Table 3. Descriptively, the scenario that elicited the highest dread risk was the terrorist attack scenario, while the fracture scenario elicited the lowest dread risk.

Table 3. *Means and standard errors for perceived risk in response to each scenario.*

Scenario	Mean	Standard Error
Depression	3.49	0.18
Fire	3.66	0.15
Fracture	2.29	0.17
Chemical Incident	3.89	0.19
Cancer	3.44	0.18
Terrorist Attack	4.07	0.15
Airplane Crash	3.80	0.18
Alcohol Dependence	3.32	0.11

Inferential analysis revealed a significant difference in perceived dread risk across scenarios, $F(7, 112) = 19.00$, $p < .001$. In order to identify whether any scenarios were, statistically (rather than descriptively), indistinguishable in perceived dread risk when compared to the descriptively high and low dread risk condition (i.e., where there are descriptive differences in mean values which are not significant), we conducted pairwise comparisons between all scenarios and: a) the high dread risk scenario (terrorist attack), and; b) the low dread risk scenario (fracture). All scenarios elicited greater dread risk than the fracture scenario ($p < .001$). In contrast, the depression ($p = .005$), fire ($p = .015$), fracture ($p < .001$) , cancer ($p < .001$), and alcohol dependence ($p < .001$) scenarios elicited less dread risk than the terrorist attack scenario (all other *p*s > .171). In short, the fracture scenario therefore elicited the least dread risk, with the terrorist attack, chemical incident, and airplane crash eliciting the joint most dread risk.

Conclusions and implications for main study scenarios. Ideally, four scenarios would have emerged which elicited a combination of high and low dread risk (i.e., one high risk/ low stigma scenario, one low risk/ high stigma scenario etc.). Although we did find that the fracture scenario elicited both the least stigma and the least dread risk (Low/ Low), there was less clarity for the remaining ideal combinations (e.g., the high stigma scenario did not elicit high dread risk). Given this, we opted to create composite scenarios combining elements from the highest and lowest dread risk/ stigma scenarios for the main study. Given the outcomes of the pilot test, we selected physical health (fracture) vs. mental health (alcohol dependence) as the stigma component, and terror attack vs routine healthcare (alcoholism/ fracture) as the risk component to vary across scenarios.

10.2 Appendix 2

Final scenarios used for main study.

Low Threat Risk/ Low Stigma
Imagine that you have fallen down the stairs and have fractured your ankle. You visited the Doctor and fortunately it is a clean fracture and no surgery is required. The Doctor has put your ankle in a plaster cast and you are well on the way to recovery.

Sharing information, such as demographic details and information about any treatment may provide valuable insights into effective treatment and response, as well as potentially informing future policy with regards to treatment and the allocation of funding. The following questions relate to this scenario.

Low Threat Risk/ High Stigma
Imagine that you are concerned about the amount of alcohol that you are drinking and the effects that this is having on your life. You visited your Doctor about this and they referred you to a specialist support team and you are well on the way to recovery.

Sharing information, such as demographic details and information about any treatment may provide valuable insights into effective treatment, as well as potentially informing future policy with regards to treatment and the allocation of funding. The following questions relate to this scenario.

High Threat Risk/ Low Stigma
Imagine that you have witnessed a terrorist attack, during which a lone gunman attacks a crowd of people. As a result of the incident, you have fallen and have fractured your ankle. You visited the Doctor and fortunately it is a clean fracture and no surgery is required. The Doctor has put your ankle is in a plaster cast and you are well on the way to recovery.

Sharing information, such as demographic details and information about any treatment may provide valuable insights into effective treatment, as well as potentially informing future policy with regards to treatment and the allocation of funding. The following questions relate to this scenario.

High Threat Risk/ High Stigma
Imagine that you have witnessed a terrorist attack, during which a lone gunman attacks a crowd of people. You have not been physically hurt, but you witnessed the injury of several other people. Since this, you have been drinking more and are concerned about the amount of alcohol that you are drinking and the effects that this is having on your life. You visited your doctor about this and they referred you to a specialist support team and you are well on the way to recovery.

Sharing information, such as demographic details and information about any treatment may provide valuable insights into effective treatment, as well as potentially informing future policy with regards to treatment and the allocation of funding. The following questions relate to this scenario.

10.3 Appendix 3

Factor loading tables for outcome measures.

The pattern matrix for some items (Table 1) are not reported since they all load on a single factor.

Table 1. *Items that load on single factors*

Risk	Stigma	Data Level
High	High	Anonymised
High	High	Linked-identifiable
High	Low	Identifiable
Low	High	Linked-identifiable
Low	High	Identifiable
Low	Low	Anonymised
Low	Low	Linked-identifiable

Low	Low	Identifiable

Table 2. *Pattern matrix for questionnaire items in the high risk and high stigma scenario with identifiable data.*

Factor	1	2
I would be happy for my data to be shared with my doctor.	.928	
I would be happy for my data to be shared with academic researchers from a university.	.619	.427
I would be happy for my data to be shared with a state healthcare organisation, e.g. NHS.	.932	
I would be happy for my data to be shared with a healthcare related government organisation, e.g. Public Health England.	.907	
I would be happy for my data to be shared with a non-healthcare related government organisation.	.513	
I would be happy for my data to be shared with an independent charity, e.g. Cancer Research UK.		.881
I would be happy for my data to be shared with a private medical research company.		.947

Table 3. *Pattern matrix for questionnaire items in the high risk and low stigma scenario with anonymised data.*

Factor	1	2
I would be happy for my data to be shared with my doctor.	.977	

I would be happy for my data to be shared with academic researchers from a university.	.725	
I would be happy for my data to be shared with a state healthcare organisation, e.g. NHS.	.951	
I would be happy for my data to be shared with a healthcare related government organisation, e.g. Public Health England.	.872	
I would be happy for my data to be shared with a non-healthcare related government organisation.		.624
I would be happy for my data to be shared with an independent charity, e.g. Cancer Research UK.		.864
I would be happy for my data to be shared with a private medical research company.		.910

Table 4. *Pattern matrix for questionnaire items in the high risk and low stigma scenario with linked-identifiable data.*

Factor	1	2
I would be happy for my data to be shared with my doctor.	.925	
I would be happy for my data to be shared with academic researchers from a university.	.838	
I would be happy for my data to be shared with a state healthcare organisation, e.g. NHS.	.961	
I would be happy for my data to be shared with a healthcare related government organisation, e.g. Public Health England.	.900	
I would be happy for my data to be shared with a non-healthcare related government organisation.	.598	

I would be happy for my data to be shared with an independent charity, e.g. Cancer Research UK.		.828
I would be happy for my data to be shared with a private medical research company.		.909

Table 5. *Pattern matrix for questionnaire items in the low risk and high stigma scenario with anonymised data.*

Factor	1	2
I would be happy for my data to be shared with my doctor.	.907	
I would be happy for my data to be shared with academic researchers from a university.	.814	
I would be happy for my data to be shared with a state healthcare organisation, e.g. NHS.	.952	
I would be happy for my data to be shared with a healthcare related government organisation, e.g. Public Health England.	.925	
I would be happy for my data to be shared with a non-healthcare related government organisation.		.649
I would be happy for my data to be shared with an independent charity, e.g. Cancer Research UK.		.838
I would be happy for my data to be shared with a private medical research company.		.873